

The rotation group in plate tectonics and the representation of uncertainties of plate reconstructions

T. Chang¹ J. Stock² and P. Molnar³

¹Department of Mathematics, University of Virginia, Charlottesville, Virginia 22903-3199, USA

²Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138, USA

³Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Accepted 1989 December 27. Received 1989 December 27; in original form 1989 August 11

SUMMARY

The calculation of the uncertainty in an estimated rotation requires a parametrization of the rotation group; that is, a unique mapping of the rotation group to a point in 3-D Euclidean space, R^3 . Numerous parametrizations of a rotation exist, including: (1) the latitude and longitude of the axis of rotation and the angle of rotation; (2) a representation as a Cartesian vector with length equal to the rotation angle and direction parallel to the rotation axis; (3) Euler angles; or (4) unit length quaternions (or, equivalently, Cayley–Klein parameters).

The uncertainty in a rotation is determined by the effect of nearby rotations on the rotated data. The uncertainty in a rotation is small, if rotations close to the best fitting rotation degrade the fit of the data by a large amount, and it is large, if only rotations differing by a large amount cause such a degradation. Ideally, we would like to parametrize the rotations in such a way so that their representation as points in R^3 would have the property that the distance between two points in R^3 reflects the effects of the corresponding rotations on the fit of the data. It can be shown mathematically that this is impossible, but for rotations of small angle, it can be done to close approximation by using vectors in Cartesian coordinates. Thus, we are led to parametrizing the uncertainty separately from the parametrization of the best fitting rotation. This approach results in simpler, more efficient calculations than if uncertainties are described in terms of rotation parameters (i.e., latitude, longitude, and the angle). We illustrate this with the example of equations for determining the uncertainty in a composite rotation from the uncertainties of its constituents.

Key words: plate tectonics, rotation.

1 INTRODUCTION

Reduced to its essentials, plate tectonics is the quantitative geometric description of the evolution of the lithosphere, in which the relative movements and relative positions of lithospheric plates are described by finite rotations on the sphere about axes passing through the centre of the Earth. Most geophysicists employ the same mental image for such a rotation: the orientation of the axis is specified by the latitude and longitude of the intersection of the axis with the Earth's surface, and the angle of rotation about the axis is the third quantity that uniquely describes the rotation. One of our purposes in this paper is to argue that although the description of finite rotations in terms of these three quantities (latitude, longitude, angle) provides a simple visual image of the rotation and is useful for many

situations, it is not appropriate for representing and combining the uncertainties in rotations.

The essential operation of a rotation is matrix multiplication. A rotation can be described as a 3×3 matrix, with nine components, subject to six constraints: that the columns be mutually orthogonal, satisfy the right-hand rule, and that each column has a length of 1. Thus although containing nine numbers, the rotation can be uniquely specified by three values. Various parametrizations of a rotation in terms of three independent values exist. In plate tectonics the most common representation is the latitude and longitude of the rotation axis on the Earth, and the angle of rotation (λ, θ, ρ). Other parametrizations are: three Euler angles (α, β, γ) (e.g. Altmann 1986); the Cartesian coordinates of a vector of length equal to the rotation angle, and oriented parallel to the axis of rotation

[\mathbf{t} or (t_1, t_2, t_3) , a column vector in 3-D]; and unit quaternions [\mathbf{q} or (q_0, q_1, q_2, q_3) —a unit column vector in four-space, or four values limited by one constraint]. Although the mathematics of deriving the nine components of the rotation matrix needed for a rotation differs according to the parametrization, any of these parametrizations can be used to obtain that matrix (see the Appendix).

In this paper we address the question of how best to describe uncertainties in a rotation. Clearly it is possible to express the uncertainties in terms of uncertainties in the nine components of the rotation matrix, or in terms of uncertainties in any of the other sets of three independent parameters describing the rotation. We attempt to show that the mathematical complexity of such a description varies greatly with choice of parametrization, but that the mathematically simplest parametrization is preferable in the sense that it yields the most accurate representation of the uncertainty.

To make our point, we attempt to draw on three levels of sophistication. First, we examine vectors and their parametrization as an analogue that illustrates difficulties with parametrization. Then we consider some simple aspects of rotations and argue heuristically that a representation of their uncertainties should utilize a different image from that given by the uncertainties in the individual parameters that describe it. Finally, we exploit some of the basics of group theory and differential geometry to show that the properties of the rotation group require such an approach. This three-tiered presentation clearly will be redundant for some readers. As a mathematician and two earth scientists, however, we have found such redundancies necessary for communication among us, and hence we anticipate that many other readers would face the same dilemma.

2 THE PARAMETRIZATION OF VECTORS AND THEIR UNCERTAINTIES

All readers should share the same understanding of a 3-D vector $\mathbf{r} = (x, y, z)$ as a line from the origin to a point in space defined by three Cartesian coordinates (x, y , and z), three polar coordinates (colatitude φ , longitude θ , length ρ) or three cylindrical coordinates (r, θ, z). Two vectors \mathbf{r}_1 and \mathbf{r}_2 differ by another vector, \mathbf{r}_3 . If \mathbf{r}_3 is written in Cartesian coordinates, it is simply given by the differences between the components: $(x_3, y_3, z_3) = (x_2 - x_1, y_2 - y_1, z_2 - z_1)$. In polar coordinates, the difference between two vectors $\mathbf{r}_2 - \mathbf{r}_1$ is easily visualized, but except for very special cases, the vector representing this difference is not equal to the differences between the components; that is, $(\varphi_3, \theta_3, \rho_3) \neq (\varphi_2 - \varphi_1, \theta_2 - \theta_1, \rho_2 - \rho_1)$. Rather, \mathbf{r}_3 must be rewritten into polar coordinates yielding

$$\begin{aligned}\varphi_3 &= \cos^{-1} \left(\frac{\rho_2 \cos \varphi_2 - \rho_1 \cos \varphi_1}{\rho_3} \right), \\ \theta_3 &= \tan^{-1} \left(\frac{\rho_2 \sin \varphi_2 \sin \theta_2 - \rho_1 \sin \varphi_1 \sin \theta_1}{\rho_2 \sin \varphi_2 \cos \theta_2 - \rho_1 \sin \varphi_1 \cos \theta_1} \right), \\ \rho_3^2 &= \rho_1^2 + \rho_2^2 - 2\rho_1\rho_2(\sin \varphi_1 \sin \varphi_2 \cos \theta_1 \cos \theta_2 \\ &\quad + \sin \varphi_1 \sin \varphi_2 \sin \theta_1 \sin \theta_2 + \cos \varphi_1 \cos \varphi_2).\end{aligned}$$

Calculation of the difference between these two vectors in

polar coordinates is laborious because the physical distance spanned by the vector $(\Delta\varphi, 0, 0)$ depends on $\rho(ds = \rho d\varphi)$ and the physical distance spanned by $(0, \Delta\theta, 0)$ depends on ρ and φ ($ds = \rho \sin \varphi d\theta$). Similar problems occur with the use of cylindrical coordinates. Thus, for calculations involving the differences between vectors, Cartesian coordinates lead to the simplest mathematics.

2.1 Parametrization of uncertainties in vectors

Basically, the question of the uncertainty in a vector reduces to a notion of distance, or the difference between two vectors. Suppose \mathbf{r} , a single vector representing an observation, has a 3-D Gaussian, or normal, probability distribution with mean vector μ_0 . The uncertainty of \mathbf{r} as a measurement of μ_0 can be described by a 'confidence set'—a family of Cartesian vectors μ close to \mathbf{r} in Cartesian space, which is interpreted as the collection of μ that are possible candidates for μ_0 , given the observation \mathbf{r} . This family of vectors can also be described by another family of vectors $\Delta\mathbf{r}$ that can be added to \mathbf{r} to yield a range of possible or plausible vectors: $\mu = \mathbf{r} + \Delta\mathbf{r}$. Although all such vectors can be represented with numerous parametrizations (using for instance, Cartesian coordinates, polar coordinates, cylindrical coordinates, etc.) for estimating uncertainties, Cartesian coordinates are simpler, and preferable, to the others. To see this let us examine the structure of the uncertainty.

2.1.1 The covariance matrix

The uncertainty $\Delta\mathbf{r}$ can be represented as a closed surface, enveloping a family of vectors centred on the null vector (the origin) (Fig. 1). When $\Delta\mathbf{r}$ is added to \mathbf{r} , the closed surface is translated so that it is centred on the endpoint of \mathbf{r} . Because we have assumed that the vectors within $\Delta\mathbf{r}$

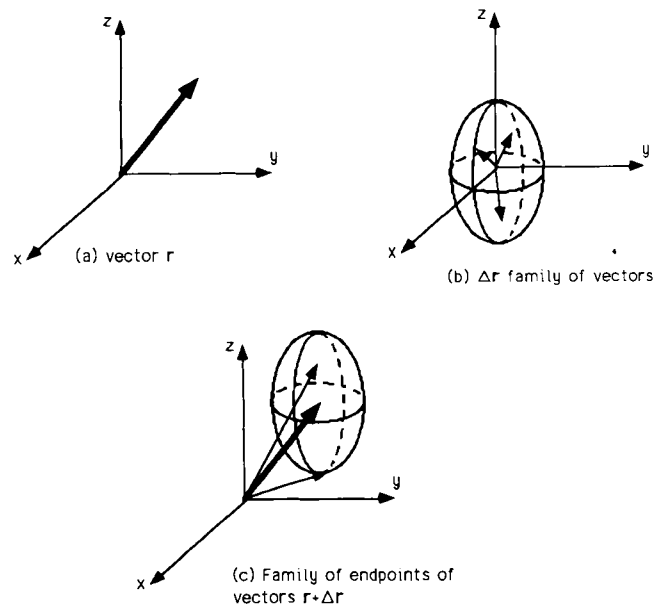


Figure 1. Representation of a vector \mathbf{r} (a) and its uncertainty region (b) as a closed surface enveloping a family $\Delta\mathbf{r}$ of vectors centred on the null vector. Resulting family of vectors (c) is obtained by translating (b) to the endpoint of \mathbf{r} .

follow a Gaussian probability distribution, the enveloping surface is an ellipsoid that can be parametrized by a matrix of variances and covariances, known as the 'covariance matrix':

$$\mathbf{M} = \text{cov}(\mathbf{r}) = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_z^2 \end{pmatrix} = E[(\mathbf{r} - \boldsymbol{\mu})(\mathbf{r} - \boldsymbol{\mu})^T]. \quad (1)$$

Here, σ_x^2 , σ_y^2 , and σ_z^2 are the variances in the x , y , and z components of \mathbf{r} , and σ_{xy} , σ_{xz} , and σ_{yz} are the covariances between each pair of components ($x-y$, $x-z$, or $y-z$) of the uncertainty (where $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$, and ρ_{ij} is the correlation coefficient). The last equality of (1) is a formal definition of \mathbf{M} as the expected value of the random 3×3 matrix $[(\Delta\mathbf{r})(\Delta\mathbf{r})^T]$. The inverse of the covariance matrix defines an ellipsoidal confidence region of the endpoints of all vectors $\boldsymbol{\mu}$ satisfying

$$(\boldsymbol{\mu} - \mathbf{r})^T \mathbf{M}^{-1} (\boldsymbol{\mu} - \mathbf{r}) = c^2 \quad (2)$$

where different values of c represent loci of the probability density of the vector \mathbf{r} . For a Gaussian distribution, c^2 is obtained from a χ^2 table with three degrees of freedom and thus $c = 2.795$ is 95 per cent confidence and $c = 1$ is 19.9 per cent confidence.

The most important aspect of this description is that $(\boldsymbol{\mu} - \mathbf{r}) = (\Delta\mathbf{r})$ enters naturally into the equation; the covariance matrix does not physically describe the family of composite vectors $\boldsymbol{\mu}$, but rather the family of small vectors $\Delta\mathbf{r}$. Therefore, if $\boldsymbol{\mu}_0$ is decreased in length to $\boldsymbol{\mu}_0/2$, but the family $\Delta\mathbf{r}$ is kept constant, then the covariance matrix \mathbf{M} will also be constant.

Finally, notice that if we add two independent random vectors \mathbf{r}_1 and \mathbf{r}_2 the covariance matrix describing the uncertainty in the resulting vector $\mathbf{r}_3 = \mathbf{r}_1 + \mathbf{r}_2$ is given simply by the sum of the covariance matrices:

$$\text{cov}(\mathbf{r}_1 + \mathbf{r}_2) = \text{cov}(\mathbf{r}_1) + \text{cov}(\mathbf{r}_2). \quad (3)$$

If \mathbf{r}_1 and \mathbf{r}_2 are not independent an additional term $2\text{cov}(\mathbf{r}_1, \mathbf{r}_2)$ must be added to the right-hand side of (3).

2.1.2 The confidence ellipsoid

By equation (2), the diagonal covariance matrix

$$\mathbf{M} = \begin{pmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix}$$

corresponds to the ellipsoid

$$\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} + \frac{z^2}{\sigma_z^2} = 1 \quad (4)$$

(since $c^2 = 1$, this is an ellipsoid of 19.9 per cent confidence). This ellipsoid has semimajor axes of length σ_x along vector direction $(1, 0, 0)$ (the x -axis), length σ_y along $(0, 1, 0)$, and length σ_z along $(0, 0, 1)$. The eigenvectors of the covariance matrix represent the directions of the major axes of the ellipsoid, and the eigenvalues (σ_x^2 , σ_y^2 , and σ_z^2) are the squares of the lengths of the semimajor axes of the ellipsoid (Fig. 2). When the off-diagonal terms of \mathbf{M} are non-zero, these properties are not as apparent by visual

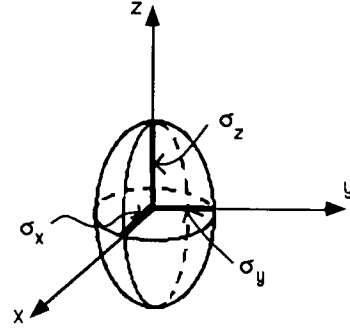


Figure 2. Ellipsoidal confidence region given by equation (4).

inspection, but it is still true that the eigenvectors and eigenvalues specify the directions and lengths of the principal axes of the ellipsoid. In general, the covariance matrix can be constructed from the matrix with column eigenvectors \mathbf{S} and the diagonal matrix of eigenvalues $\boldsymbol{\Lambda}$ by the familiar linear algebraic composition

$$\mathbf{M} = \mathbf{S}\boldsymbol{\Lambda}\mathbf{S}^{-1}. \quad (5)$$

2.2 Uncertainties when vectors are rotated and combined

The covariance matrix of a vector \mathbf{r} will change if the vector and its uncertainty ellipsoid are both rotated by a rotation $\boldsymbol{\Omega}$. The ellipsoidal volume of small vectors $\Delta\mathbf{r}$ attached to the end of \mathbf{r} undergoes the same rotation $\boldsymbol{\Omega}$ as \mathbf{r} . This corresponds to rotating the eigenvectors of the covariance matrix (the major axes of the ellipsoid) by $\boldsymbol{\Omega}$. Because the covariance matrix can be described by (5), the rotated covariance matrix \mathbf{M}' corresponds to

$$\mathbf{M}' = (\boldsymbol{\Omega}\mathbf{S})\boldsymbol{\Lambda}(\boldsymbol{\Omega}\mathbf{S})^{-1} = \boldsymbol{\Omega}\mathbf{S}\boldsymbol{\Lambda}\mathbf{S}^{-1}\boldsymbol{\Omega}^{-1} = \boldsymbol{\Omega}\mathbf{M}\boldsymbol{\Omega}^{-1}. \quad (6)$$

In other words, if a vector \mathbf{r} is rotated, its covariance matrix is transformed by pre-multiplication (post-rotation) by the rotation matrix and post-multiplication (pre-rotation) by its inverse.

If two vectors parametrized by Cartesian coordinates are given in different reference frames (a common occurrence in geophysics when points on different plates are considered), one vector with its covariance matrix must be rotated into the reference frame of the other one before the vectors or their uncertainties can be added together. Let the vector \mathbf{u} have the representation \mathbf{u}_A in reference frame A, and \mathbf{u}_B in reference frame B. Let the rotation $\boldsymbol{\Omega}$ take reference frame A to reference frame B, that is $\mathbf{u}_B = \boldsymbol{\Omega}\mathbf{u}_A$. The sum of two vectors \mathbf{u} and \mathbf{v} (in reference frame B) is

$$(\mathbf{u} + \mathbf{v})_B = \mathbf{u}_B + \mathbf{v}_B = \mathbf{u}_B + \boldsymbol{\Omega}\mathbf{v}_A. \quad (7a)$$

Recall that the rotation $\boldsymbol{\Omega}^{-1}$ takes reference frame B to reference frame A. Therefore, in reference frame A, the sum is

$$(\mathbf{u} + \mathbf{v})_A = \mathbf{u}_A + \mathbf{v}_A = \boldsymbol{\Omega}^{-1}\mathbf{u}_B + \mathbf{v}_A. \quad (7b)$$

Suppose now \mathbf{u} and \mathbf{v} are independent random vectors, with \mathbf{u} expressed in the frame B and \mathbf{v} in the frame A. The covariance matrix, in reference frame B, of their sum by (6) is

$$\text{cov}(\mathbf{u} + \mathbf{v})_B = \text{cov}(\mathbf{u}_B) + \boldsymbol{\Omega} \text{cov}(\mathbf{v}_A) \boldsymbol{\Omega}^{-1} \quad (8a)$$

and in reference frame A it is

$$\text{cov}(\mathbf{u} + \mathbf{v})_A = \mathbf{\Omega}^{-1} \text{cov}(\mathbf{u}_B) \mathbf{\Omega} + \text{cov}(\mathbf{v}_A). \quad (8b)$$

2.3 Uncertainties of combined vectors in non-Cartesian coordinate systems

If the vectors are not expressed by Cartesian coordinates, but by some other representation, their combination, and the description of their uncertainties, become more complicated than equations (7) and (8). For instance, in many parametrizations, the null vector corresponds to a singularity. A vector of zero length in φ - θ - ρ space is uniquely described by $\rho = 0$, but an infinite number of possible values can be assigned to φ and θ . Thus, any family of vectors that includes the null vector cannot be conveniently represented by a covariance matrix in (φ, θ, ρ) coordinates.

One can determine a covariance matrix for \mathbf{r} in terms of $\Delta\varphi$, $\Delta\theta$ and $\Delta\rho$, but that matrix will vary if \mathbf{r} changes. Two vectors \mathbf{r}_1 and \mathbf{r}_2 with the same covariance matrix \mathbf{M} in Cartesian coordinates will have different covariance matrices in φ - θ - ρ space, unless ρ and φ are the same. Moreover, the covariance matrix of the sum of these two vectors in φ - θ - ρ coordinates will not be the sum of their individual covariance matrices in φ - θ - ρ coordinates. This is clear if one considers two unequal vectors \mathbf{r}_S and \mathbf{r}_L , each of which is uncertain by the same amount in the sense that their covariance matrices, \mathbf{M}_S and \mathbf{M}_L , expressed in Cartesian coordinates, are identical. Let us assume that $\rho_L \gg \rho_S$. If $\|\Delta\mathbf{r}\| \ll \rho_L$, the family of vectors $\mathbf{r}_L + \Delta\mathbf{r}$ points in nearly the same direction as \mathbf{r}_L ; hence $\Delta\theta$ must be small. If ρ_S is comparable with $\|\Delta\mathbf{r}\|$, the ellipsoid described by the family of vectors $\mathbf{r}_S + \Delta\mathbf{r}$ will span a larger range of θ and φ

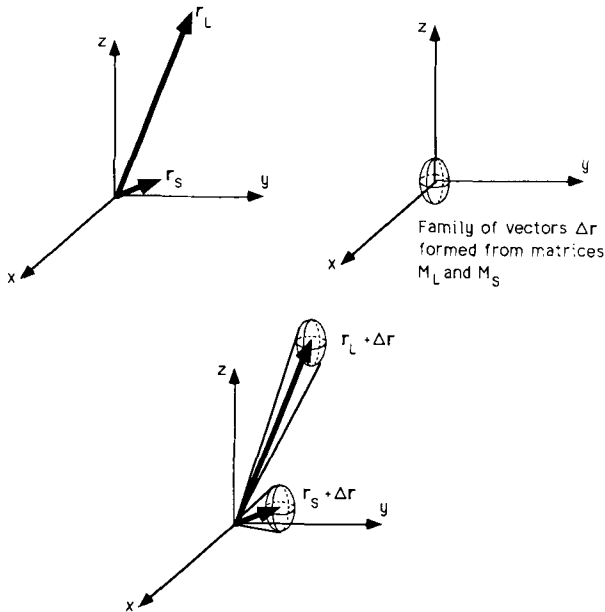


Figure 3. Distortion of the uncertainty of a vector parametrized by spherical coordinates. The same covariance matrix in Cartesian coordinates ($\mathbf{M}_S = \mathbf{M}_L$) defines the sizes of uncertainty regions for short and long vectors \mathbf{r}_S and \mathbf{r}_L . If the uncertainty regions were described in spherical coordinates, the range of θ and φ values would be much greater for $\mathbf{r}_S + \Delta\mathbf{r}$ than for $\mathbf{r}_L + \Delta\mathbf{r}$.

(Fig. 3). Thus $\Delta\theta$ and $\Delta\varphi$ will be large, and although the same uncertainty, $\Delta\mathbf{r}$, was assigned to both \mathbf{r}_S and \mathbf{r}_L , the uncertainties in \mathbf{r}_S and \mathbf{r}_L expressed as covariance matrices in coordinates (φ, θ, ρ) are very different.

For the purposes of combining uncertainties, it is clearly preferable for the covariance matrices of a parameter to be independent of the value of the parameter. Thus, for a vector, the representation of its uncertainties in Cartesian coordinates is preferable to a polar coordinate representation; the Cartesian representation leads to the simplest mathematics when these uncertainties are combined.

Expressing the uncertainty of a vector in terms of uncertainties in its polar coordinates is also inaccurate. Consider a vector pointing in the x -direction: $\mathbf{r} = (x_1, 0, 0)$ in Cartesian coordinates, or $\mathbf{r} = (\pi/2, 0, x_1)$ in polar coordinates (φ, θ, ρ) . Imagine that only the uncertainty in the y -direction is important: $\Delta\mathbf{r} = (0, \pm\Delta y, 0)$ in Cartesian coordinates. Clearly $\Delta\varphi = 0$, and $\Delta\theta = \arctan(\Delta y/x_1)$. Notice that it is difficult to express the uncertainty in the polar coordinate value of ρ , which ranges from values of x_1 (when $y = 0$) to $\sqrt{x_1^2 + (\Delta y)^2}$ (when $y = \pm\Delta y$). Clearly one cannot write $\rho \pm \Delta\rho$ and define $\Delta\rho$ sensibly. Thus any description of the uncertainty in \mathbf{r} in terms of uncertainties in ρ , θ , and φ is likely to distort that uncertainty.

In summary, for the purposes of combining two vectors and their uncertainties, the representation of these vectors by their Cartesian coordinates is simplest, because (a) the sum of the two vectors is easily written in terms of the sum of their components; (b) the covariance matrix transforms simply under rotation; and (c) the final covariance matrix is the sum of the individual covariance matrices. For other representation of vectors (polar coordinates, direction cosines, etc.) additional transformations are required before the components of the vector or of its associated covariance matrix can be summed.

3 DIFFICULTIES IN PARAMETRIZING UNCERTAINTIES IN ROTATIONS

Mathematically a rotation of the sphere is given by a 3×3 matrix \mathbf{R} subject to the conditions

$$\mathbf{R}\mathbf{R}^T = \mathbf{R}^T\mathbf{R} = \mathbf{I}, \quad (9)$$

$$\det \mathbf{R} = 1, \quad (10)$$

where \mathbf{I} is the 3×3 identity matrix, and \det means the determinant. Equations (9) and (10) represent six independent conditions on the nine entries of \mathbf{R} and hence the matrix \mathbf{R} is uniquely defined by three parameters. By analogy with the case of vectors described previously, we seek to choose these three parameters so that the uncertainty in \mathbf{R} is expressed in the simplest possible fashion.

3.1 How the uncertainty in a rotation is determined

To understand the role of parametrization in the expression of uncertainties in rotations, it is necessary first to understand the nature of that uncertainty. In plate tectonics, the rotation is not a quantity which can be directly observed but rather must be estimated from data. Following standard statistical notation we will use \mathbf{R} to denote the unknown

rotation and $\hat{\mathbf{R}}$ its estimate. In statistical language, the rotation \mathbf{R} is an 'unknown parameter', and its estimate $\hat{\mathbf{R}}$ is subject to error caused by the errors in the data. In this sense the situation we discuss here is different from what we discussed in the previous section on random vectors. In that section, we assumed an error distribution for \mathbf{r} as an estimate of $\boldsymbol{\mu}$; here the errors of $\hat{\mathbf{R}}$ as an estimate of \mathbf{R} must be derived from the errors in the data. Statistics studies this process: the transmission of error from the data to the estimate. Since most scientists understand the errors in their data better than the errors in their estimates, statistics provides valuable insight into the errors of the estimate.

A formal derivation of the error in $\hat{\mathbf{R}}$ is beyond the scope of this paper, but can be found in a series of papers by Chang (1986, 1988, 1989) and Rivest (1989) with various assumptions about both the types of data and the probability distributions of errors. A heuristic approach merely to bounding the uncertainties in $\hat{\mathbf{R}}$ was given by Stock & Molnar (1983; see also Molnar & Stock 1985). The basic approach of both to estimating such uncertainties is to observe how *perturbing* rotations, when combined with $\hat{\mathbf{R}}$, affect the fit of the data. *The uncertainty in an estimated rotation $\hat{\mathbf{R}}$ is large, if a second rotation with a large angle can be combined with it without significantly degrading the data that the rotation is supposed to fit. The uncertainty is small, when only small rotations about any axis can be added to $\hat{\mathbf{R}}$ without degrading the quality of the fit.*

If the estimate $\hat{\mathbf{R}}$ is written as a matrix, and its uncertainty is given by a family $\Delta\mathbf{R}$ of possible small rotations (also written as matrices), the resulting family of possible rotations is $(\Delta\mathbf{R})_{\mathbf{R}}\hat{\mathbf{R}}$, or $\hat{\mathbf{R}}(\Delta\mathbf{R})_{\mathbf{U}}$, depending upon whether the small rotations in $\Delta\mathbf{R}$ are defined in the rotated (\mathbf{R}) or unrotated (\mathbf{U}) frame of reference. In the most general case, the rotations in $\Delta\mathbf{R}$ can be small rotations about any axis, just as the vectors in $\Delta\mathbf{r}$ could be small vectors pointing in any direction.

The uncertainty in the orientation of the rotation axis (or pole position) in, for example, $\hat{\mathbf{R}}(\Delta\mathbf{R})_{\mathbf{U}}$ depends upon the angle of the rotation. If one combines a large rotation with a small one, the axis of the resultant rotation is close to that of the larger rotation, but if one combines two comparable rotations, the axis of the combination can differ greatly from the axes of the component rotations. It follows that the uncertainties in the orientations of rotation axes (or in pole positions on the Earth) are much larger for small rotations (corresponding to plate reconstructions for relatively recent geologic times), than for larger rotations, if the uncertainties in the rotations are the same. (This fact, however, is somewhat obscured in plate reconstructions because reconstructions for more recent periods are often better constrained and less uncertain.) In any case, the uncertainty in the orientation of the rotation axis of \mathbf{R} , represented by the rotations in the family $\hat{\mathbf{R}}(\Delta\mathbf{R})_{\mathbf{U}}$, can give a very misleading impression of the uncertainty in \mathbf{R} .

3.2 The effect of parametrization on the shape and size of an uncertainty region

From the above discussion, if we use lat.–long.–angle (λ, θ, ρ) to parametrize the rotation, the size of the uncertainty region in (λ, θ, ρ) -space will show great distortions depending upon the value of ρ . This

phenomenon is analogous to the distortions in $\Delta\mathbf{r}$ when the vectors in 3-space are expressed in polar coordinates. Similarly, the use of a pseudo-vector \mathbf{t} to parametrize $\hat{\mathbf{R}}$ will cause the uncertainty region to show great distortions depending upon the value of $\rho = \|\mathbf{t}\|$. Indeed, it will be seen in the next section that, if the uncertainty in $\hat{\mathbf{R}}$ is determined by the degradation of the fit of data, no single parametrization of rotations will preserve the size and shape of the uncertainty region for all $\hat{\mathbf{R}}$.

The way out of this dilemma is first to choose a parametrization that least distorts the size and shape of a region of *small rotations*, and then to use this parametrization to express the permissible $\Delta\mathbf{R}$. The rotation estimate $\hat{\mathbf{R}}$ can be calculated using any parametrization, and then its uncertainty can be expressed as $\hat{\mathbf{R}}(\Delta\mathbf{R})$. In this way, the rotations close to $\hat{\mathbf{R}}$ are parametrized by their differences from $\hat{\mathbf{R}}$. We will call such a parametrization a 'moving' parametrization because the origin of the parametrization is placed at $\hat{\mathbf{R}}$ and not at the null rotation or some other pre-determined fixed rotation.

By analogy with the example of vectors, we might expect that the form of the parametrization will affect the ease with which two rotations are combined and more importantly the ease of the calculation of the uncertainty in a combined rotation. We thus seek a parametrization in which these calculations are straightforward.

When two rotations are portrayed as matrices \mathbf{A} and \mathbf{B} , the combined rotation is portrayed by the product matrix \mathbf{AB} . If one uses a three-parameter description (lat.–long.–angle: λ, θ, ρ ; Euler angles: α, β, γ ; pseudo-vector: \mathbf{t}) of the rotations, there is no simple analogy with the summation of vectors; awkwardness in determining the combined rotation will always arise from the awkwardness of the translation from the three-parameter description to the matrix description. Thus, it should be obvious that, in general, the matrix \mathbf{AB} is not the rotation given by $\mathbf{R}(\lambda_a + \lambda_b, \theta_a + \theta_b, \rho_a + \rho_b)$, the rotation $\mathbf{R}(\alpha_a + \alpha_b, \beta_a + \beta_b, \gamma_a + \gamma_b)$, or the rotation $\mathbf{R}(\mathbf{t}_a + \mathbf{t}_b)$. If then the uncertainty in \mathbf{AB} is calculated using any of these three-parameter descriptions, the awkwardness of the translation propagates into an awkwardness of the necessary formulae.

A 'moving' parametrization eliminates this problem by eliminating the need to determine the matrix product \mathbf{AB} as an intermediate step in the calculation of its uncertainty. Rather, the uncertainty $\Delta(\mathbf{AB})$ is expressed in terms of $\hat{\mathbf{A}}, \hat{\mathbf{B}}, \Delta\mathbf{A}$, and $\Delta\mathbf{B}$, and the possible rotations are those of the form $\hat{\mathbf{A}}\hat{\mathbf{B}}[\Delta(\mathbf{AB})]$.

Since in a moving parametrization the rotations close to $\hat{\mathbf{R}}$ are parametrized by their differences from $\hat{\mathbf{R}}$, parametrization of these small differences by (λ, θ, ρ) or by (α, β, γ) is ruled out because such representations have singularities at the origin (corresponding to the null rotation).

3.3 What is the covariance matrix of a fitted rotation?

For a random p -dimensional vector \mathbf{r} with mean vector (that is, expected value) $\boldsymbol{\mu}_0$, the covariance matrix $\text{cov}(\mathbf{r})$ of \mathbf{r} is defined by equation (1) as the expected value of the $p \times p$ matrix $(\mathbf{r} - \boldsymbol{\mu}_0)(\mathbf{r} - \boldsymbol{\mu}_0)^T$. The expected value is a probabilistic concept; both $\boldsymbol{\mu}_0$ and $\text{cov}(\mathbf{r})$ depend upon the probability distribution of the random vector \mathbf{r} .

As discussed above, the probability distribution of any estimated parameter, and in particular of a fitted rotation $\hat{\mathbf{R}}$, is derived from an assumed probability of the data and the method of deriving the estimate. Thus, in particular, $\text{cov}(\hat{\mathbf{R}})$ is a derived quantity.

For a fitted rotation, another complication arises. The rotations do not comprise a linear subspace (hyperplane) in some Euclidean space. Yet, the concept of expected value (and *a fortiori* mean and covariance) requires that the fitted parameter (in this case $\hat{\mathbf{R}}$) vary over some linear subspace. Thus, strictly, $\text{cov}(\hat{\mathbf{R}})$ cannot be not defined. Rather, we must replace $\hat{\mathbf{R}}$ by some 3-D vector description $\hat{\mathbf{r}}$ of it, and use $\text{cov}(\hat{\mathbf{r}})$ as a surrogate for $\text{cov}(\hat{\mathbf{R}})$. For example if we employ the lat.-long.-angle parametrization (λ, θ, ρ) , we would use the covariance matrix of the fitted values $(\hat{\lambda}, \hat{\theta}, \hat{\rho})$.

In summary, the covariance matrix for a fitted rotation cannot be defined independently of a specific parametrization of the rotations, and the choice of the parametrization will profoundly affect the form of the covariance matrix. Any formula or calculation of such a matrix must involve a derivation of how the errors in the data are transmitted into the covariance matrix of the fitted rotation for some reasonably plausible probability model of the data.

3.4 The moving exponential parametrization

Consider the representation of the small rotations by 'pseudo-vectors': that is, if \mathbf{h} is a three-vector such that $\|\mathbf{h}\| \ll 1$, $\Phi(\mathbf{h})$ is a rotation of $\|\mathbf{h}\|$ radians around the axis $\mathbf{h}/\|\mathbf{h}\|$. This representation of the rotation matrix is called the 'exponential representation' [Appendix A; see also Altmann (1986, pp. 73–75)]. When we parametrize rotations near $\hat{\mathbf{R}}$ in the form $\hat{\mathbf{R}}\Phi(\mathbf{h})$, we shall say that we are adopting a 'moving exponential parametrization'.

In the moving exponential parametrization, the uncertainty in $\hat{\mathbf{R}}$ is expressed using the covariance matrix $\text{cov}(\hat{\mathbf{h}})$ of the three-vector $\hat{\mathbf{h}}$ which satisfies $\hat{\mathbf{R}} = \mathbf{R}\Phi(\hat{\mathbf{h}})$. Here \mathbf{R} is the true (unknown) rotation. It is readily seen that $\hat{\mathbf{R}} = \mathbf{R}\Phi(\hat{\mathbf{h}})$ is equivalent to $\mathbf{R} = \hat{\mathbf{R}}\Phi(-\hat{\mathbf{h}})$.

If we have an estimated rotation $\hat{\mathbf{R}}$, the possible rotations \mathbf{R} are then of the form $\mathbf{R} = \hat{\mathbf{R}}\Phi(\hat{\mathbf{h}})$ where $\hat{\mathbf{h}}$ satisfies the equation

$$\hat{\mathbf{h}}^T \text{cov}(\hat{\mathbf{h}})^{-1} \hat{\mathbf{h}} < c^2 \quad (11)$$

and c^2 is obtained from a χ^2 table with three degrees of freedom (or possibly an F table with three degrees of freedom in the numerator). The collection of such \mathbf{R} is said to be a confidence region for the unknown rotation. Equation (11) for rotations is analogous to equation (2) for vectors.

For an error model that can be reasonably entertained for magnetic anomaly lineation and fracture zone data, if $\hat{\mathbf{R}}$ is estimated using Hellinger's (1981) criterion, the formula for $\text{cov}(\hat{\mathbf{h}})$ and its derivation are given in Chang (1988). Covariance matrices for other least-squares estimates of $\hat{\mathbf{R}}$ for other models for the errors in data are given in Chang (1987).

The square roots of the three eigenvalues of $\text{cov}(\hat{\mathbf{h}})$ define (up to a constant of proportionality) three small angles of rotation about three orthogonal axes defined by the

corresponding eigenvectors of $\text{cov}(\hat{\mathbf{h}})$. These eigenvectors and their associated eigenvalues are therefore statistically rigorous equivalents of the Partial Uncertainty Rotations discussed by Stock & Molnar (1983; see also Molnar & Stock 1985).

Given two independent estimated rotations $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$, the covariance matrix of the combined rotation $\hat{\mathbf{C}} = \hat{\mathbf{A}}\hat{\mathbf{B}}$ is given by

$$\text{cov}(\hat{\mathbf{h}}_{\mathbf{C}}) \approx \hat{\mathbf{B}}^T \text{cov}(\hat{\mathbf{h}}_{\mathbf{A}})\hat{\mathbf{B}} + \text{cov}(\hat{\mathbf{h}}_{\mathbf{B}}). \quad (12)$$

Equation (12) depends only upon the use of a moving exponential parametrization and not upon the derivation of the covariance matrix. A proof of equation (12) is given in the Appendix. The \approx indicates that a linear approximation is used in that derivation. The approximation is that the rotations $\Phi(\hat{\mathbf{h}}_{\mathbf{A}})$ and $\Phi(\hat{\mathbf{h}}_{\mathbf{B}})$ are small enough that they can be treated as infinitesimal rotations which add as vectors; this works well if the eigenvalues of $\text{cov}(\hat{\mathbf{h}}_{\mathbf{A}})$ and $\text{cov}(\hat{\mathbf{h}}_{\mathbf{B}})$ are $\gg 1$, corresponding to $\|\hat{\mathbf{h}}_{\mathbf{A}}\|$ and $\|\hat{\mathbf{h}}_{\mathbf{B}}\| \ll 1$. Notice that (12) is much simpler than the corresponding equations [Appendix equations (A3) and (A5)] for parametrization by lat.-long.-angle (λ, θ, ρ) or pseudo-vector \mathbf{t} and is strikingly similar to equations (8).

4 THE ROTATION GROUP AND ITS GEOMETRY

In this section we discuss the structure of the collection of all possible rotation matrices, those matrices that satisfy equations (9) and (10). The standard mathematical name for this collection of matrices is the 'special orthogonal group in three dimensions', $\text{SO}(3)$.

For the purposes of calculation and especially for statistical evaluation, the equations (9) and (10) are not a very useful description of rotations. What is needed is a parametrization of $\text{SO}(3)$, a mapping of points in Euclidean 3-D space (hereafter called \mathbf{R}^3) into $\text{SO}(3)$. Equations in the Appendix (Sections A1, A2, A3) define some such mappings, or parametrizations. As we argued above, the choice of parametrization will vitally affect the quality of results, and in particular most of these parametrizations are very poor for statistical analysis. A good parametrization should yield simple mathematical formulae and be characterized by rapid convergence of the asymptotic expansions that permeate statistical computations.

The errors in the data are defined in terms of distance on the surface of the sphere. Each rotation manifests itself as a map of points on the sphere, 'the action of the rotation group on the sphere'. Suppose we have a notion of the 'distance' $D(\mathbf{A}, \mathbf{B})$ between two rotations, \mathbf{A} and \mathbf{B} such that $D(\mathbf{A}, \mathbf{B})$ reflects the distance between $\mathbf{A}\mathbf{r}$ and $\mathbf{B}\mathbf{r}$ as \mathbf{r} varies over the sphere. If this were true, $D[\hat{\mathbf{R}}, \hat{\mathbf{R}}(\Delta\mathbf{R})]$ would be small exactly when $\hat{\mathbf{R}}(\Delta\mathbf{R})$ does not significantly degrade the fit (given by $\hat{\mathbf{R}}$) of the data points. As discussed in the previous section, the statistical assessment of the error in $\hat{\mathbf{R}}$ is based upon this degradation of the fit. It follows that $D[\hat{\mathbf{R}}, \hat{\mathbf{R}}(\Delta\mathbf{R})]$ is intimately related to the error in $\hat{\mathbf{R}}$.

An obvious way to define such a distance is

$D(\mathbf{A}, \mathbf{B}) = \text{maximum distance in radians on the unit sphere from } \mathbf{A}\mathbf{r} \text{ to } \mathbf{B}\mathbf{r}, \text{ among all } \mathbf{r} \text{ on the unit sphere.}$

Since rotation by \mathbf{B}^T of both $\mathbf{A}\mathbf{r}$ and $\mathbf{B}\mathbf{r}$ preserves distance, $D(\mathbf{A}, \mathbf{B}) = D(\mathbf{B}^T\mathbf{A}, \mathbf{I})$ = maximum distance on the unit sphere from $\mathbf{B}^T\mathbf{A}\mathbf{r}$ to \mathbf{r} , for all possible \mathbf{r}
 = rotation angle of $\mathbf{B}^T\mathbf{A}$. (13)

We will justify the selection of this distance function in another way: using the notion of 'invariance'. After doing so, we will contend that any parametrization of $\text{SO}(3)$ will distort $D(\mathbf{A}, \mathbf{B})$ somewhere on $\text{SO}(3)$, but what the moving exponential parametrization has minimal distortion in the region of interest: rotations close to $\hat{\mathbf{R}}$. This accounts for the simplicity of results, such as equation (12), derived using a moving exponential parametrization.

4.1 Invariance

Rotations form a mathematical 'group' because (a) there exists an operation by which rotations can be combined to form new rotations and (b) each rotation has an inverse. The first of these properties simply means that if we perform rotation \mathbf{A} and then rotation \mathbf{B} , the combination is a third rotation \mathbf{C} given by $\mathbf{C} = \mathbf{B}\mathbf{A}$, the product of the 3×3 matrices corresponding to \mathbf{A} and \mathbf{B} . The second simply means that for each rotation \mathbf{A} , there exists an inverse \mathbf{A}^{-1} , such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. In rotations the transpose of \mathbf{A} , \mathbf{A}^T , is the inverse: $\mathbf{A}^T = \mathbf{A}^{-1}$.

The most obvious type of invariance is invariance under change of basis. Suppose (x, y, z) is a point on the sphere and \mathbf{C} is a matrix whose rows form a right-hand rule orthonormal basis of Euclidean space. Then if (x', y', z') are the coordinates of (x, y, z) expressed in terms of the basis consisting of the rows of \mathbf{C} , it follows that

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \mathbf{C} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

\mathbf{C} is a matrix of direction cosines relating (x, y, z) to (x', y', z') . If \mathbf{A} is a rotation matrix that moves a rigid body on the sphere defined by (x, y, z) from one place to another, then in the coordinate system (x', y', z') the operation that moves the same rigid body to the same place is $\mathbf{C}\mathbf{A}\mathbf{C}^T$. Presumably statistical inferences should not depend upon the basis, or coordinate system, that is used. For such an invariance under a change of coordinates, if data \mathbf{r}_i on the sphere are changed to $\mathbf{C}\mathbf{r}_i$, then inferences about \mathbf{A} become inferences about $\mathbf{C}\mathbf{A}\mathbf{C}^T$. Because the rows of \mathbf{C} form a right-hand orthonormal basis of \mathbb{R}^3 , the matrix \mathbf{C} will also lie in $\text{SO}(3)$. For a definition of distance to meet this requirement of invariance,

$$D(\mathbf{C}\mathbf{A}\mathbf{C}^T, \mathbf{C}\mathbf{B}\mathbf{C}^T) = D(\mathbf{A}, \mathbf{B}). \quad (14)$$

Let us consider a second, more subtle form of invariance. Suppose we have a collection of points \mathbf{r}_i , which are part of a rigid plate, and a rotation \mathbf{C} representing a reconstruction of the plate defined by \mathbf{r}_i at some time t_1 . Now consider a second rotation \mathbf{A} that describes the rotation of points $\mathbf{C}\mathbf{r}_i$, representing the reconstruction of the plate from its position at t_1 to its position at an earlier time t_2 . $\mathbf{A}\mathbf{C}$ would then describe the rotation of \mathbf{r}_i to their positions at t_2 . Assuming \mathbf{C} is known without error, then 'invariance' implies that inferences about $\mathbf{A}\mathbf{C}$ become inferences about \mathbf{A} . Thus

$$D(\mathbf{A}\mathbf{C}, \mathbf{B}\mathbf{C}) = D(\mathbf{A}, \mathbf{B}). \quad (15)$$

The distance function (13) meets these two requirements. $D(\mathbf{C}\mathbf{A}\mathbf{C}^T, \mathbf{C}\mathbf{B}\mathbf{C}^T)$ = angle of rotation of $(\mathbf{C}\mathbf{B}\mathbf{C}^T)^T(\mathbf{C}\mathbf{A}\mathbf{C}^T) = \mathbf{C}\mathbf{B}^T\mathbf{C}^T\mathbf{C}\mathbf{A}^T = \mathbf{C}\mathbf{B}^T\mathbf{A}^T$, which is simply the matrix for the rotation $\mathbf{B}^T\mathbf{A}$ with respect to the basis consisting of the rows of \mathbf{C} . $\mathbf{C}\mathbf{B}^T\mathbf{A}^T$ is the same rotation as $\mathbf{B}^T\mathbf{A}$, but in a different coordinate system; its angle of rotation is the same. The proof of (15) is similar: $D(\mathbf{A}\mathbf{C}, \mathbf{B}\mathbf{C})$ = angle of rotation of $(\mathbf{B}\mathbf{C})^T(\mathbf{A}\mathbf{C}) = \mathbf{C}^T\mathbf{B}^T\mathbf{A}\mathbf{C}$ and the latter matrix is the rotation $\mathbf{B}^T\mathbf{A}$ with respect to the basis consisting of the rows of \mathbf{C}^T .

In the Appendix part A6, we show that the distance function (13) is the *only* distance function $D(\mathbf{A}, \mathbf{B})$ which satisfies (14) and (15).

4.2 Geometry of the rotation group

We seek a 'picture' of the rotation group $\text{SO}(3)$ in which the distance function (13) is faithfully represented. Such a picture is best provided by unit quaternions: the collection \mathbf{Q} of \mathbf{q} of the form

$$\mathbf{q} = q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k}, \quad (16)$$

$$q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1.$$

A good reference on the basic properties of quaternions as a representation is Altmann (1986). Some of these properties are given in Appendix, Part A5. Part A5 also describes a mathematically equivalent representation in terms of Cayley–Klein parameters. We wish to emphasize the following properties of the quaternion representation.

(i) Each unit length quaternion \mathbf{q} can be written in the form

$$\mathbf{q} = \cos(\rho/2) + \sin(\rho/2)\mathbf{u} \quad (17)$$

where \mathbf{u} is a unit length pure quaternion of the form $u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}$. Using standard engineering notation, \mathbf{u} can be thought of as a unit length vector on the sphere (in Euclidean 3-space). Then \mathbf{q} represents right-hand rule rotation of ρ radians around the axis \mathbf{u} . Let $\mathbf{A}(\mathbf{q})$ be the matrix for this rotation.

(ii) It follows that each rotation can be represented by two quaternions: if \mathbf{q} is one representation, then $-\mathbf{q}$ is the other [in symbols, $\mathbf{A}(\mathbf{q}) = \mathbf{A}(-\mathbf{q})$]. Mathematicians say that \mathbf{Q} *doubly covers* $\text{SO}(3)$. In particular the null rotation is represented by the quaternions $1 (= 1 + 0\mathbf{i} + 0\mathbf{j} + 0\mathbf{k})$ and -1 [$\mathbf{A}(1) = \mathbf{A}(-1) = \mathbf{I}$].

(iii) If $\bar{\mathbf{q}} = q_0 - q_1\mathbf{i} - q_2\mathbf{j} - q_3\mathbf{k}$, then $\mathbf{A}(\bar{\mathbf{q}}) = \mathbf{A}(\mathbf{q})^T$, and $\mathbf{A}(\bar{\mathbf{q}})$ and $\mathbf{A}(\mathbf{q})$ are inverses of each other. Furthermore $\mathbf{A}(\mathbf{q}_1\mathbf{q}_2) = \mathbf{A}(\mathbf{q}_1)\mathbf{A}(\mathbf{q}_2)$.

(iv) From (16), we see that \mathbf{Q} is a sphere of radius 1 in Euclidean 4-D space. Let $d(\mathbf{q}_1, \mathbf{q}_2)$ be the surface distance on \mathbf{Q} between \mathbf{q}_1 and \mathbf{q}_2 , defined as the arc cosine of the dot product $\mathbf{q}_1 \cdot \mathbf{q}_2$. As shown in the Appendix, part A6, for any $\mathbf{q}, \mathbf{q}_1, \mathbf{q}_2$

$$d(\mathbf{q}\mathbf{q}_1, \mathbf{q}\mathbf{q}_2) = d(\mathbf{q}_1, \mathbf{q}_2)$$

and

$$d(\mathbf{q}_1\mathbf{q}, \mathbf{q}_2\mathbf{q}) = d(\mathbf{q}_1, \mathbf{q}_2).$$

What is the relationship between D and d ? Thinking of 1 as the 4-D analogue of the North Pole, the distance $d(\mathbf{q}, 1)$ of \mathbf{q} to 1 is the colatitude of \mathbf{q} . If we write \mathbf{q} in the form (17), $d(\mathbf{q}, 1) = \rho/2$. Thus if the quaternion \mathbf{q} has a

non-negative real part, we have

$$D[\mathbf{A}(\mathbf{q}), \mathbf{I}] = 2d(\mathbf{q}, \mathbf{I}).$$

We would like to hypothesize that

$$D[\mathbf{A}(\mathbf{q}_1), \mathbf{A}(\mathbf{q}_2)] = 2d(\mathbf{q}_1, \mathbf{q}_2).$$

Because of the double covering property (ii), this is not quite true. Instead we have

$$D[\mathbf{A}(\mathbf{q}_1), \mathbf{A}(\mathbf{q}_2)] = 2 \min [d(\mathbf{q}_1, \mathbf{q}_2), d(\mathbf{q}_1, -\mathbf{q}_2)]. \quad (18)$$

Proof of equation (18):

$$\begin{aligned} D[\mathbf{A}(\mathbf{q}_1), \mathbf{A}(\mathbf{q}_2)] &= D[\mathbf{A}(\mathbf{q}_1)\mathbf{A}(\mathbf{q}_2)^T, \mathbf{I}], \quad [\text{using (15)}] \\ &= D[\mathbf{A}(\mathbf{q}_1\tilde{\mathbf{q}}_2), \mathbf{I}], \quad [\text{using (iii)}] \\ &= \text{either } 2d(\mathbf{q}_1\tilde{\mathbf{q}}_2, \mathbf{I}), \quad (\text{if real part of } \mathbf{q}_1\tilde{\mathbf{q}}_2 \geq 0) \\ &\quad \text{or } 2d(-\mathbf{q}_1\tilde{\mathbf{q}}_2, \mathbf{I}), \quad (\text{otherwise}) \\ &= 2 \min [d(\mathbf{q}_1\tilde{\mathbf{q}}_2, \mathbf{I}), d(-\mathbf{q}_1\tilde{\mathbf{q}}_2, \mathbf{I})] \\ &= 2 \min [d(\mathbf{q}_1, \mathbf{q}_2), d(\mathbf{q}_1, -\mathbf{q}_2)], \quad [\text{using (iv)}]. \end{aligned}$$

The moral of this section is: *The rotation group SO(3) has an intrinsic and essentially unique invariant distance function; this distance function is relevant to the study of rigid body motions; and this distance function can be visualized (except for an annoying double covering) by Euclidean distance on the unit sphere Q in Euclidean 4-D space.*

4.3 The effect of parametrization on the apparent size and shape of a confidence region

A parametrization of SO(3) is essentially a mapping $f: \mathbf{R}^3 \rightarrow \text{SO}(3)$ in that each point of SO(3) is associated by f to a point of \mathbf{R}^3 . A confidence region for \mathbf{R} is a region $\mathbf{C}(\hat{\mathbf{R}})$ in SO(3) consisting of rotations of the form $\hat{\mathbf{R}}(\Delta\mathbf{R})$ for small $\Delta\mathbf{R}$. Ideally $f^{-1}[\mathbf{C}(\hat{\mathbf{R}})]$, the inverse image of $\mathbf{C}(\hat{\mathbf{R}})$ in \mathbf{R}^3 , should have a shape and a size accurately reflecting its shape and size in SO(3).

'Shape and size' are properties derived from the notion of distance. We have seen that SO(3) has a special distance function, given by equation (13), such that $D[\hat{\mathbf{R}}(\Delta\mathbf{R}), \hat{\mathbf{R}}]$ measures the degradation by $\hat{\mathbf{R}}(\Delta\mathbf{R})$ of the fit provided by $\hat{\mathbf{R}}$. Thus what we seek in a parametrization f is that it take standard Euclidean distance in \mathbf{R}^3 into the distance D in SO(3):

$$\|\mathbf{x} - \mathbf{y}\| = D[f(\mathbf{x}), f(\mathbf{y})]$$

for points \mathbf{x} and \mathbf{y} in \mathbf{R}^3 .

The geometry of SO(3) with the distance function D is close to the geometry of \mathbf{Q} , a unit sphere in \mathbf{R}^4 , with Euclidean distance on the surface of that sphere. A distance preserving parametrization $f: \mathbf{R}^3 \rightarrow \text{SO}(3)$ cannot exist because of the curvature of \mathbf{Q} , for exactly the same mathematical reason (but in one higher dimension) that shape- and size-preserving mappings of the Earth's surface (that is $f: \mathbf{R}^2 \rightarrow \text{Earth}$) do not exist. Nevertheless given a specific small region of interest, f can be chosen to preserve reasonably faithfully the distance relationships in that region of interest. Thus we use a parametrization centred on $\hat{\mathbf{R}}$; we parametrize rotations close to $\hat{\mathbf{R}}$ in the form $\hat{\mathbf{R}}f(\mathbf{x})$ for \mathbf{x} in \mathbf{R}^3 with $\|\mathbf{x}\| \ll 1$.

Among the possible parametrizations of small rotations, one wants to choose a parametrization f that does not greatly distort small rotations. A good choice is $f = \Phi$, the exponential parametrization, leading finally to a moving exponential parametrization, described above.

If a good parametrization is chosen, our formulae (for example, to describe the errors in $\hat{\mathbf{R}}$) become simpler, because a poor parametrization introduces distortions in the region of interest (the rotations close to $\hat{\mathbf{R}}$). Attempts to compensate for these distortions lead to complications in theorems and formulae. The Appendix contains analogues (A3) and (A5) to formula (12) for the lat.-long.-angle and (fixed) exponential parametrizations, and it can be seen that the moving exponential parametrization provides by far the simplest results.

5 CONCLUSION

The use of unit quaternions allows one to visualize important geometric properties of rotations that are relevant to determining the errors in estimates of rotations that describe the reconstructions of lithospheric plates. Any parametrization will distort these geometric relationships. Nevertheless, a moving exponential parametrization (Section 3.4) centred near the true rotation \mathbf{A} distorts these relationships least in the region of interest: small perturbations of \mathbf{A} . The moving exponential parametrization also results in invariant inferences about \mathbf{A} .

A distorting parametrization complicates calculations since the calculations will reflect the corrections necessary to compensate for the distortions introduced by the parametrization. It follows that a moving exponential parametrization is the preferred parametrization for calculating close to \mathbf{A} . In addition, since it would be astonishing if nature allowed a distortion to improve the quality of asymptotic approximations, asymptotic approximations are most likely to be of high quality if a moving exponential parametrization is used.

ACKNOWLEDGMENTS

The first named author was funded in part by NSF grant #DMS-8901494, the second by NSF grant #OCE-8911809, and the third by NASA grant #NAG-795.

REFERENCES

- Altmann, S., 1986. *Rotations, Quaternions and Double Groups*, Oxford University Press, Oxford, UK.
- Chang, T., 1986. Spherical regression, *Ann. Stat.*, **14**, 907–924.
- Chang, T., 1987. On the statistical properties of estimated rotations, *J. geophys. Res.*, **92B**, 6319–6329.
- Chang, T., 1988. On reconstructing tectonic plate motion from ship track crossings, *J. Am. Stat. Assoc.*, **83**, 1178–1183.
- Chang, T., 1989. Spherical regression with errors in variables, *Ann. Stat.*, **17**, 293–306.
- Goldstein, H., 1950. *Classical Mechanics*, Addison-Wesley, Reading, MA.
- Hellinger, S. J., 1981. The uncertainties of finite rotations in plate tectonics, *J. geophys. Res.*, **86B**, 9312–9318.
- Jurdy, D. M. & Stefanick, M., 1987. Errors in plate rotations as described by covariance matrices and their combination in reconstructions, *J. geophys. Res.*, **92B**, 6310–6318.
- Marion, J. B., 1970. *Classical Dynamics of Particles and Systems*, Academic Press, New York.

- Molnar, P. & Stock, J., 1985. A method for bounding uncertainties in combined plate reconstructions, *J. geophys. Res.*, **90**, 12 537–12 544.
- Rivest, L., 1989. Spherical regression for concentrated Fisher-von Mises distributions *Ann. Stat.*, **17**, 307–317.
- Stock, J. & Molnar, P., 1983. Some geometrical aspects of uncertainties in combined plate reconstructions, *Geology*, **11**, 697–701.

APPENDIX: PARAMETRIZATIONS OF ROTATIONS

A1 Exponential parametrization ('pseudo-vectors')

Here we will always use \mathbf{t} to denote a 3×1 vector whose coordinates are (t_1, t_2, t_3) . Let $\Phi(\mathbf{t})$ be the right-hand rule rotation of $\rho = \|\mathbf{t}\| = (t_1^2 + t_2^2 + t_3^2)^{1/2}$ radians around the axis $\mathbf{t}/\|\mathbf{t}\|$. Letting

$$\mathbf{T} = \begin{pmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{pmatrix},$$

it follows that $\Phi(\mathbf{t}) = \sum_{r=0}^{\infty} \mathbf{T}^r / r! = e^{\mathbf{T}}$ (e.g. Altmann 1986, pp. 73–75). This sum reduces to

$$\Phi(\mathbf{t}) = \mathbf{I} + \frac{\sin \rho}{\rho} \mathbf{T} + \frac{1 - \cos \rho}{\rho^2} \mathbf{T}^2. \quad (\text{A1})$$

A1.1 Combining two rotations

Let \mathbf{t}_a , \mathbf{t}_b , and \mathbf{t}_c be defined by $\Phi(\mathbf{t}_a) = \mathbf{A}$, $\Phi(\mathbf{t}_b) = \mathbf{B}$, and $\Phi(\mathbf{t}_c) = \mathbf{C}$ respectively. Let $\rho_a = \|\mathbf{t}_a\|$ and $\mathbf{h}_a = \mathbf{t}_a / \|\mathbf{t}_a\|$ and define similar quantities for \mathbf{B} and \mathbf{C} . If $\mathbf{C} = \mathbf{AB}$, we have

$$\begin{aligned} \cos \frac{\rho_c}{2} &= \cos \frac{\rho_a}{2} \cos \frac{\rho_b}{2} - \sin \frac{\rho_a}{2} \sin \frac{\rho_b}{2} (\mathbf{h}_a \cdot \mathbf{h}_b), \\ \sin \frac{\rho_c}{2} \mathbf{h}_c &= \sin \frac{\rho_a}{2} \cos \frac{\rho_b}{2} \mathbf{h}_a + \cos \frac{\rho_a}{2} \sin \frac{\rho_b}{2} \mathbf{h}_b \\ &\quad + \sin \frac{\rho_a}{2} \sin \frac{\rho_b}{2} (\mathbf{h}_a \times \mathbf{h}_b), \end{aligned} \quad (\text{A2})$$

where $\mathbf{h}_a \cdot \mathbf{h}_b$ and $\mathbf{h}_a \times \mathbf{h}_b$ represent the scalar and vector products of \mathbf{h}_a and \mathbf{h}_b .

A1.2 Combining two covariance matrices

If the exponential parametrization is used, covariance matrices can be combined using a formula given by Jurdy & Stefanick (1987). We reproduce their result, modifying only the notation to be consistent with that used here. Let $\hat{\mathbf{t}}_a$ be defined by $\Phi(\hat{\mathbf{t}}_a) = \hat{\mathbf{A}}$ and similarly define $\hat{\mathbf{t}}_b$ and $\hat{\mathbf{t}}_c$ where $\hat{\mathbf{C}} = \hat{\mathbf{A}}\hat{\mathbf{B}}$. Furthermore define vectors \mathbf{a} , \mathbf{b} , \mathbf{c} , $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, and $\hat{\mathbf{c}}$ by

$$\mathbf{a} = \frac{\sin \|\hat{\mathbf{t}}_a/2\|}{\|\hat{\mathbf{t}}_a\|} \hat{\mathbf{t}}_a$$

and similarly for the rest.

$$\text{cov}(\hat{\mathbf{c}}) \approx \mathbf{G} \text{cov}(\hat{\mathbf{a}}) \mathbf{G}^T + \mathbf{F} \text{cov}(\hat{\mathbf{b}}) \mathbf{F}^T. \quad (\text{A3})$$

Here \approx means a linear approximation has been used and the 3×3 matrices \mathbf{F} and \mathbf{G} are defined by

$$F_{ij} = \delta_{ij} \sqrt{1 - \|\hat{\mathbf{a}}\|^2} - \frac{\hat{a}_i \hat{a}_j}{\sqrt{1 - \|\hat{\mathbf{a}}\|^2}} - \varepsilon_{ijk} \hat{a}_k,$$

$$G_{ij} = \delta_{ij} \sqrt{1 - \|\hat{\mathbf{b}}\|^2} - \frac{\hat{b}_i \hat{b}_j}{\sqrt{1 - \|\hat{\mathbf{b}}\|^2}} + \varepsilon_{ijk} \hat{b}_k$$

where \hat{a}_i and \hat{b}_i are the components of $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ respectively, δ is the delta function, and ε_{ijk} is the alternating tensor with $\varepsilon_{123} = 1$ (and hence $\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = -\varepsilon_{132} = -\varepsilon_{321} = -\varepsilon_{213} = 1$, all others = 0). The repeated index summation convention is used.

Equation (A3) actually calculates the covariance matrix of the 'squashed exponential parameter' $\hat{\mathbf{c}}$, but the covariance matrix of the exponential parameter $\hat{\mathbf{t}}_c$ can be derived from it. We recall that how $\text{cov}(\hat{\mathbf{a}})$ or $\text{cov}(\hat{\mathbf{t}}_a)$ is actually defined or calculated depends upon the method of estimating $\hat{\mathbf{A}}$ and the assumed error model for the data and that $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ are assumed independent.

A2 Parametrization by axis latitude, longitude and angle of rotation

We will continue to use λ , θ , and ρ to denote the axis latitude, axis longitude, and angle of rotation, all written in radians and will use $\Psi(\lambda, \theta, \rho)$ to denote the corresponding rotation. To express $\Psi(\lambda, \theta, \rho)$ as a matrix, we use (A1)

$$\Psi(\lambda, \theta, \rho) = \Phi(\mathbf{t})$$

where

$$\mathbf{t} = \rho \begin{pmatrix} \cos \lambda \cos \theta \\ \cos \lambda \sin \theta \\ \sin \lambda \end{pmatrix}. \quad (\text{A4})$$

A2.1 Combining rotations

This is also best done via the exponential parametrization. Given $\mathbf{A} = \Psi(\lambda_a, \theta_a, \rho_a)$ and $\mathbf{B} = \Psi(\lambda_b, \theta_b, \rho_b)$ we can use (A4) to define \mathbf{t}_a and \mathbf{t}_b . Using (A2), we can calculate $\mathbf{t}_c = (t_{1c}, t_{2c}, t_{3c})$ and

$$\begin{aligned} \rho_c &= \|\mathbf{t}_c\|, \\ \sin \lambda_c &= t_{3c} / \rho_c, \\ \tan \theta_c &= t_{2c} / t_{1c}. \end{aligned}$$

A2.2 Combining covariance matrices

Letting $\hat{\mathbf{A}} = \Psi(\hat{\lambda}_a, \hat{\theta}_a, \hat{\rho}_a)$ and similarly for $\hat{\mathbf{B}}$ and $\hat{\mathbf{C}}$, we have

$$\mathbf{D} \text{cov}(\hat{\lambda}_c, \hat{\theta}_c, \hat{\rho}_c) \mathbf{D}^T \approx \mathbf{G} \text{cov}(\hat{\lambda}_a, \hat{\theta}_a, \hat{\rho}_a) \mathbf{G}^T + \mathbf{F} \text{cov}(\hat{\lambda}_b, \hat{\theta}_b, \hat{\rho}_b) \mathbf{F}^T \quad (\text{A5})$$

where the 3×3 matrices \mathbf{D} , \mathbf{G} , and \mathbf{F} are defined by

$$D_{11} = -\sin \frac{\hat{\rho}_c}{2} \cos \hat{\theta}_c \sin \hat{\lambda}_c,$$

$$D_{12} = -\sin \frac{\hat{\rho}_c}{2} \sin \hat{\theta}_c \cos \hat{\lambda}_c,$$

$$D_{13} = \frac{1}{2} \cos \frac{\hat{\rho}_c}{2} \cos \hat{\theta}_c \cos \hat{\lambda}_c,$$

$$D_{22} = \sin \frac{\hat{\rho}_c}{2} \cos \hat{\theta}_c,$$

$$D_{23} = \frac{1}{2} \cos \frac{\hat{\rho}_c}{2} \sin \hat{\theta}_c,$$

$$D_{33} = -\frac{1}{2} \sin \frac{\hat{\rho}_c}{2},$$

$$D_{21} = D_{31} = D_{32} = 0,$$

$$G_{11} = -\sin \frac{\hat{\rho}_a}{2} \cos \frac{\hat{\rho}_b}{2} \cos \hat{\theta}_a \sin \hat{\lambda}_a \\ + \sin \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \cos \hat{\theta}_a \sin \hat{\theta}_b \cos \hat{\lambda}_a,$$

$$F_{11} = -\cos \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \cos \hat{\theta}_b \sin \hat{\lambda}_b \\ - \sin \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \sin \hat{\theta}_a \cos \hat{\theta}_b \cos \hat{\lambda}_b,$$

$$G_{21} = -\sin \frac{\hat{\rho}_a}{2} \cos \frac{\hat{\rho}_b}{2} \sin \hat{\theta}_a \cos \hat{\lambda}_a + \sin \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \\ \times (-\sin \hat{\theta}_a \sin \hat{\theta}_b \sin \hat{\lambda}_a - \cos \hat{\theta}_a \cos \hat{\theta}_b \sin \hat{\lambda}_b),$$

$$F_{21} = -\cos \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \sin \hat{\theta}_b \cos \hat{\lambda}_b + \sin \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \\ \times (\cos \hat{\theta}_a \cos \hat{\theta}_b \sin \hat{\lambda}_a + \sin \hat{\theta}_a \sin \hat{\theta}_b \sin \hat{\lambda}_b),$$

$$2G_{31} = \cos \frac{\hat{\rho}_a}{2} \cos \frac{\hat{\rho}_b}{2} \cos \hat{\theta}_a \cos \hat{\lambda}_a - \sin \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \cos \hat{\theta}_b \cos \hat{\lambda}_b \\ + \cos \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \\ \times (\cos \hat{\theta}_a \sin \hat{\theta}_b \sin \hat{\lambda}_a - \sin \hat{\theta}_a \cos \hat{\theta}_b \sin \hat{\lambda}_b),$$

$$2F_{31} = -\sin \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \cos \hat{\theta}_a \cos \hat{\lambda}_a + \cos \frac{\hat{\rho}_a}{2} \cos \frac{\hat{\rho}_b}{2} \cos \hat{\theta}_b \cos \hat{\lambda}_b \\ + \sin \frac{\hat{\rho}_a}{2} \cos \frac{\hat{\rho}_b}{2} \\ \times (\cos \hat{\theta}_a \sin \hat{\theta}_b \sin \hat{\lambda}_a - \sin \hat{\theta}_a \cos \hat{\theta}_b \sin \hat{\lambda}_b),$$

$$G_{12} = -F_{12} = -\sin \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \cos \hat{\theta}_a \cos \hat{\theta}_b \cos (\hat{\lambda}_b - \hat{\lambda}_a),$$

$$G_{22} = \sin \frac{\hat{\rho}_a}{2} \cos \frac{\hat{\rho}_b}{2} \cos \hat{\theta}_a \\ - \sin \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \sin \hat{\theta}_a \cos \hat{\theta}_b \sin (\hat{\lambda}_b - \hat{\lambda}_a),$$

$$F_{22} = \cos \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \cos \hat{\theta}_b \\ - \sin \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \cos \hat{\theta}_a \sin \hat{\theta}_b \sin (\hat{\lambda}_b - \hat{\lambda}_a),$$

$$2G_{32} = \cos \frac{\hat{\rho}_a}{2} \cos \frac{\hat{\rho}_b}{2} \sin \hat{\theta}_a - \sin \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \sin \hat{\theta}_b \\ + \cos \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \cos \hat{\theta}_a \cos \hat{\theta}_b \sin (\hat{\lambda}_b - \hat{\lambda}_a),$$

$$2F_{32} = -\sin \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \sin \hat{\theta}_a + \cos \frac{\hat{\rho}_a}{2} \cos \frac{\hat{\rho}_b}{2} \sin \hat{\theta}_b \\ + \sin \frac{\hat{\rho}_a}{2} \cos \frac{\hat{\rho}_b}{2} \cos \hat{\theta}_a \cos \hat{\theta}_b \sin (\hat{\lambda}_b - \hat{\lambda}_a),$$

$$G_{13} = -F_{13} = -\sin \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \cos \hat{\theta}_a \cos \hat{\theta}_b \sin (\hat{\lambda}_b - \hat{\lambda}_a),$$

$$G_{23} = -\sin \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \\ \times [-\sin \hat{\theta}_a \cos \hat{\theta}_b \cos (\hat{\lambda}_b - \hat{\lambda}_a) + \cos \hat{\theta}_a \sin \hat{\theta}_b],$$

$$F_{23} = -\sin \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \\ \times [-\cos \hat{\theta}_a \sin \hat{\theta}_b \cos (\hat{\lambda}_b - \hat{\lambda}_a) + \sin \hat{\theta}_a \cos \hat{\theta}_b],$$

$$2G_{33} = -\sin \frac{\hat{\rho}_a}{2} \cos \frac{\hat{\rho}_b}{2} - \cos \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} \\ \times [\cos \hat{\theta}_a \cos \hat{\theta}_b \cos (\hat{\lambda}_b - \hat{\lambda}_a) + \sin \hat{\theta}_a \sin \hat{\theta}_b],$$

$$2F_{33} = -\cos \frac{\hat{\rho}_a}{2} \sin \frac{\hat{\rho}_b}{2} - \sin \frac{\hat{\rho}_a}{2} \cos \frac{\hat{\rho}_b}{2} \\ \times [\cos \hat{\theta}_a \cos \hat{\theta}_b \cos (\hat{\lambda}_b - \hat{\lambda}_a) + \sin \hat{\theta}_a \sin \hat{\theta}_b].$$

The proof of the formula (A5) is necessarily extremely messy and will not be given here.

A3 Definition of Euler Angles

A rotation can be specified as three successive rotations about orthogonal axes. The angles of these rotations are known as Euler angles and are taken to be positive counterclockwise. There are numerous different conventions regarding the order of axes used in the rotations (see Goldstein 1950; Marion 1970; Altmann 1986). Below we adopt the convention advocated by Altmann (1986) that the axes \mathbf{x} , \mathbf{y} , and \mathbf{z} are fixed in space (relative to the rotating body) and that the rotations are applied in the following order:

first, $\Phi(\gamma\mathbf{z})$ = rotation by γ around \mathbf{z} ,

second, $\Phi(\beta\mathbf{y})$ = rotation by β around \mathbf{y} ,

third, $\Phi(\alpha\mathbf{z})$ = rotation by α around \mathbf{z} .

With this convention, the matrix of the total rotation $\mathbf{E}(\alpha, \beta, \gamma)$ is equal to

$$\mathbf{E}(\alpha, \beta, \gamma) = \Phi(\alpha\mathbf{z})\Phi(\beta\mathbf{y})\Phi(\gamma\mathbf{z}) \\ = \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{pmatrix} \\ \times \begin{pmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ = \begin{pmatrix} \cos \alpha \cos \beta \cos \gamma - \sin \alpha \sin \gamma & -\cos \alpha \cos \beta \sin \gamma - \sin \alpha \cos \gamma & \cos \alpha \sin \beta \\ \sin \alpha \cos \beta \cos \gamma + \cos \alpha \sin \gamma & -\sin \alpha \cos \beta \sin \gamma + \cos \alpha \cos \gamma & \sin \alpha \sin \beta \\ -\sin \beta \cos \gamma & \sin \beta \sin \gamma & \cos \beta \end{pmatrix}$$

A4 Proof of equation (12): covariance of combined rotations in the moving exponential parametrization

Let \mathbf{A} , \mathbf{B} , and $\mathbf{C} = \mathbf{AB}$, be the true (unknown) matrices and write $\hat{\mathbf{A}} = \mathbf{A}\Phi(\hat{\mathbf{h}}_a)$, $\hat{\mathbf{B}} = \mathbf{B}\Phi(\hat{\mathbf{h}}_b)$, and $\hat{\mathbf{C}} = \mathbf{C}\Phi(\hat{\mathbf{h}}_c)$, where $\hat{\mathbf{C}} = \hat{\mathbf{A}}\hat{\mathbf{B}}$. Then

$$\mathbf{AB}\Phi(\hat{\mathbf{h}}_c) = \hat{\mathbf{C}} = \hat{\mathbf{A}}\hat{\mathbf{B}} = \mathbf{A}\Phi(\hat{\mathbf{h}}_a)\mathbf{B}\Phi(\hat{\mathbf{h}}_b),$$

and hence

$$\Phi(\hat{\mathbf{h}}_c) = \mathbf{B}^T\Phi(\hat{\mathbf{h}}_a)\mathbf{B}\Phi(\hat{\mathbf{h}}_b)$$

Recall that Φ has the properties

$$\Phi(\mathbf{Bt}) = \mathbf{B}\Phi(\mathbf{t})\mathbf{B}^T$$

and $\Phi(\mathbf{t} + \mathbf{s}) = \Phi(\mathbf{t})\Phi(\mathbf{s}) + \text{terms of quadratic order in } \|\mathbf{t}\| \text{ and } \|\mathbf{s}\|$. Because $\|\hat{\mathbf{h}}_c\|$, $\|\hat{\mathbf{h}}_a\|$, and $\|\hat{\mathbf{h}}_b\|$ are small, it follows that

$$\hat{\mathbf{h}}_c \approx \mathbf{B}^T \hat{\mathbf{h}}_a + \hat{\mathbf{h}}_b. \quad (\text{A6})$$

From (3) and (A6)

$$\text{cov}(\hat{\mathbf{h}}_c) \approx \mathbf{B}^T \text{cov}(\hat{\mathbf{h}}_a) \mathbf{B} + \text{cov}(\hat{\mathbf{h}}_b). \quad (\text{A7})$$

(A7) expresses $\text{cov}(\hat{\mathbf{h}}_c)$ in terms of the unknown true matrix \mathbf{B} . To estimate $\text{cov}(\hat{\mathbf{h}}_c)$ it is necessary to further approximate by replacing \mathbf{B} with its estimate $\hat{\mathbf{B}}$. The result is equation (12).

Although more abstract, the proof of (12) is simpler than the proofs of the formulae (A3) and (A5), which use the exponential and lat.–long.–angle parametrizations, respectively. Moreover, (12) is algebraically much simpler than (A3) and (A5) and hence is less prone to calculation error.

A5 Quaternions and Cayley–Klein parameters

The quaternions are an extension of the complex numbers in a manner analogous to the definition of the complex numbers as an extension of the real numbers. More formally a quaternion \mathbf{j} is arbitrarily defined with the property that $\mathbf{j}^2 = -1$ and a quaternion is defined to be an object of the form

$$\mathbf{q} = (q_0 + q_1\mathbf{i}) + (q_2 + q_3\mathbf{j}). \quad (\text{A8})$$

Letting $\mathbf{k} = \mathbf{ij}$, (A8) can be rewritten in the form $\mathbf{q} = q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k}$. It is seen that the quaternions can be thought of geometrically as 4-D vectors. If q_0 is zero, \mathbf{q} is said to be a *pure quaternion*.

Addition of quaternions is defined in accordance with the law of vector addition:

$$\begin{aligned} (q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k}) + (r_0 + r_1\mathbf{i} + r_2\mathbf{j} + r_3\mathbf{k}) \\ = (q_0 + r_0) + (q_1 + r_1)\mathbf{i} + (q_2 + r_2)\mathbf{j} + (q_3 + r_3)\mathbf{k}. \end{aligned}$$

Multiplication is defined using the equations

$$\begin{aligned} \mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1, \quad \mathbf{ij} = -\mathbf{ji} = \mathbf{k}, \\ \mathbf{ki} = -\mathbf{ik} = \mathbf{j}, \quad \mathbf{jk} = -\mathbf{kj} = \mathbf{i} \end{aligned} \quad (\text{A9})$$

and the distributive laws

$$(\mathbf{q} + \mathbf{r})\mathbf{s} = \mathbf{qs} + \mathbf{rs} \quad \text{and} \quad \mathbf{q}(\mathbf{r} + \mathbf{s}) = \mathbf{qr} + \mathbf{qs}.$$

If this is done, the quaternions satisfy all the usual rules of algebra *except* the commutative law for multiplication: that is, \mathbf{qr} is usually not equal to \mathbf{rq} . It is in fact impossible to define a multiplication on 4-D vectors which will satisfy all the usual rules of algebra including the commutative law.

The *conjugate* of the quaternion $\mathbf{q} = q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k}$ is defined to be $\bar{\mathbf{q}} = q_0 - q_1\mathbf{i} - q_2\mathbf{j} - q_3\mathbf{k}$. Note that $\mathbf{q}\bar{\mathbf{q}} = \bar{\mathbf{q}}\mathbf{q}$.

Let \mathbf{Q} be the collection of *unit length quaternions*, that is quaternions which satisfy $q_0^2 + q_1^2 + q_2^2 + q_3^2 = \mathbf{q}\bar{\mathbf{q}} = 1$. Such a unit length quaternion can be written in the form $\mathbf{q} = \cos(\rho/2) + \sin(\rho/2)\mathbf{u}$ where \mathbf{u} is a unit length pure quaternion. Using standard engineering notation, \mathbf{u} can be

thought of as a unit length vector on the sphere (in Euclidean 3-space). Let $\mathbf{A}(\mathbf{q})$ be the matrix representation of right-hand rule rotation of ρ radians around the axis \mathbf{u} . We first state two technical lemmas about \mathbf{Q} and $\mathbf{A}(\mathbf{q})$ which together will imply properties (i)–(iii) in Section 4.2. The proofs of these lemmas conclude this section of the appendix.

Lemma 1. Let \mathbf{q} be a unit length quaternion and let \mathbf{t} be a point in \mathbb{R}^3 . Then

$$\mathbf{A}(\mathbf{q})\mathbf{t} = \mathbf{qt}\bar{\mathbf{q}}$$

where, on the right-hand side, \mathbf{t} is considered as a pure quaternion (using standard engineering notation for vectors) and quaternion multiplication is used. As a matrix,

$$\mathbf{A}(\mathbf{q}) = \begin{pmatrix} q_1^2 + q_0^2 - q_2^2 - q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_3q_1 + q_0q_2) \\ 2(q_1q_2 + q_0q_3) & q_2^2 + q_0^2 - q_1^2 - q_3^2 & 2(q_2q_3 - q_0q_1) \\ 2(q_3q_1 - q_0q_2) & 2(q_2q_3 + q_0q_1) & q_3^2 + q_0^2 - q_1^2 - q_2^2 \end{pmatrix}. \quad (\text{A10})$$

Lemma 2. $\mathbf{A}(\mathbf{q}_1\mathbf{q}_2) = \mathbf{A}(\mathbf{q}_1)\mathbf{A}(\mathbf{q}_2)$.

A mathematically equivalent description of \mathbf{Q} and $\mathbf{A}(\mathbf{q})$ is in terms of *Cayley–Klein parameters*. Let \mathbf{I} , σ_x , σ_y , and σ_z be the *Pauli spin matrices* defined to be

$$\begin{aligned} \mathbf{I} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \\ \sigma_y &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned}$$

Note that σ_x , σ_y , and σ_z satisfy the familiar identities

$$\begin{aligned} (-i\sigma_x)^2 &= (-i\sigma_y)^2 = (-i\sigma_z)^2 = -\mathbf{I}, \\ (-i\sigma_x)(-i\sigma_y) &= -(-i\sigma_y)(-i\sigma_x) = (-i\sigma_z), \\ (-i\sigma_y)(-i\sigma_z) &= -(-i\sigma_z)(-i\sigma_y) = (-i\sigma_x), \\ (-i\sigma_z)(-i\sigma_x) &= -(-i\sigma_x)(-i\sigma_z) = (-i\sigma_y). \end{aligned} \quad (\text{A11})$$

The equations (A11) have the identical form to equations (A9). It follows that if for $\mathbf{q} = q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k}$, we define $\mathbf{H}(\mathbf{q})$ to be the 2×2 complex matrix

$$\begin{aligned} \mathbf{H}(\mathbf{q}) &= q_0 + q_1(-i\sigma_x) + q_2(-i\sigma_y) + q_3(-i\sigma_z) \\ &= \begin{pmatrix} q_0 - q_3i & -q_1i - q_2 \\ -q_1i + q_2 & q_0 + q_3i \end{pmatrix}, \end{aligned} \quad (\text{A12})$$

we have

$$\mathbf{H}(\mathbf{q})\mathbf{H}(\mathbf{r}) = \mathbf{H}(\mathbf{qr}). \quad (\text{A13})$$

Equation (A12) implies that if \mathbf{q} has unit length

$$\det \mathbf{H}(\mathbf{q}) = 1$$

and

$$\mathbf{H}(\mathbf{q})\overline{\mathbf{H}(\mathbf{q})}^T = \mathbf{I},$$

where $\bar{\mathbf{A}}^T$ denotes the conjugate transpose of the matrix \mathbf{A} . The collection of matrices \mathbf{A} satisfying $\det \mathbf{A} = 1$ and $\mathbf{A}\bar{\mathbf{A}}^T = \mathbf{I}$

is called the *special unitary group* $SU(2)$, and the map \mathbf{H} defines a one-to-one correspondence of \mathbf{Q} and $SU(2)$.

Finally, Lemma 1 and equation (A13) imply that the matrix \mathbf{A} in $SU(2)$ corresponds to the rotation which takes the 3-vector \mathbf{t} into the 3-vector \mathbf{t}' where the components of \mathbf{t} and \mathbf{t}' are related by

$$-i \begin{pmatrix} t'_3 & t'_1 - it'_2 \\ t'_1 + it'_2 & -t'_3 \end{pmatrix} = \mathbf{A} \begin{pmatrix} t_3 & t_1 - it_2 \\ t_1 + it_2 & -t_3 \end{pmatrix} \bar{\mathbf{A}}^T. \quad (\text{A14})$$

The parametrization of $SO(3)$ using $SU(2)$ and (A14) is referred to as Cayley–Klein parameters. Using the map \mathbf{H} , $SU(2)$ with the parametrization of $SO(3)$ given by (A14) is equivalent to \mathbf{Q} and the parametrization given by Lemma 1.

Proof of the Lemma 1. Let $\bar{\mathbf{A}}(\mathbf{q})$ be the matrix defined by $\bar{\mathbf{A}}(\mathbf{q}) \mathbf{t} = \mathbf{q}^T \bar{\mathbf{q}}$. It is a straightforward computation to check that $\bar{\mathbf{A}}(\mathbf{q})$ is given by the right-hand side of (A10).

Substituting

$$q_0 = \cos \frac{\rho}{2}, \quad q_1 = \sin \frac{\rho}{2} u_1,$$

$$q_2 = \sin \frac{\rho}{2} u_2, \quad q_3 = \sin \frac{\rho}{2} u_3$$

into (A10) and applying (A1) yields $\bar{\mathbf{A}}(\mathbf{q}) = \Phi[(\rho u_1, \rho u_2, \rho u_3)] = \mathbf{A}(\mathbf{q})$.

Proof of the Lemma 2. Let \mathbf{t} be a point in \mathbb{R}^3 , thought of as a pure quaternion. Then $\mathbf{A}(\mathbf{q}_1 \mathbf{q}_2) \mathbf{t} = \mathbf{q}_1 \mathbf{q}_2 \bar{\mathbf{t}} \bar{\mathbf{q}}_1 = \mathbf{q}_1 [\mathbf{A}(\mathbf{q}_2) \mathbf{t}] \bar{\mathbf{q}}_1 = \mathbf{A}(\mathbf{q}_1) \mathbf{A}(\mathbf{q}_2) \mathbf{t}$.

A6 Discussion of the uniqueness of the distance function D on $SO(3)$

In mathematics, there are several approaches to distance. One, the *metric space* approach, simply defines a distance function to be any function with certain properties. In this sense, invariant distance on $SO(3)$ is in fact not unique.

In our case, we are interested in parametrizing $SO(3)$, and then for the purpose of statistical calculations, differentiating in the resulting Euclidean parameters. It follows that the relevant notion of distance is the Riemannian definition of distance, which defines the notion of distance compatibly with a differential structure.

Thinking of $SO(3)$ as a subset of Euclidean 9-D space \mathbb{R}^9 , we can use the ordinary Euclidean definition of the derivative to define the derivative $\alpha'(t)$ of a curve $\alpha(t)$ whose image lies in $SO(3)$. Then $\alpha'(t)$ will be a vector tangent to $SO(3)$ at $\alpha(t)$, for each t . In this way we construct the *tangent vectors* to $SO(3)$. A *Riemannian metric* on $SO(3)$ is then a scalar product $\langle \mathbf{v}, \mathbf{w} \rangle$ for vectors \mathbf{v} and \mathbf{w} tangent to $SO(3)$. An example, by no means unique, of a Riemannian metric is the ordinary dot product that $SO(3)$ inherits as a subset of \mathbb{R}^9 . Given a Riemannian metric, we can define the length of a curve $\alpha(t)$ from $t=0$ to 1 using the usual integral from calculus

$$\text{Length}[\alpha(t)] = \int_{t=0}^{t=1} \langle \alpha'(t), \alpha'(t) \rangle^{1/2} dt.$$

The Riemannian distance $\bar{D}(\mathbf{A}, \mathbf{B})$ between the rotations \mathbf{A} and \mathbf{B} is then defined to be the length of the shortest curve $\alpha(t)$ with $\alpha(0) = \mathbf{A}$ and $\alpha(1) = \mathbf{B}$.

For example consider the unit quaternions \mathbf{Q} as points on the unit sphere in \mathbb{R}^4 . A tangent vector to \mathbf{Q} at the quaternion \mathbf{q} is a vector whose dot product with \mathbf{q} is zero. The ordinary Euclidean dot product in \mathbb{R}^4 defines a Riemannian metric on \mathbf{Q} , which in turn defines a Riemannian distance function on \mathbf{Q} that is the usual surface distance $d(\mathbf{q}_1, \mathbf{q}_2)$. In other words, $d(\mathbf{q}_1, \mathbf{q}_2)$ is the (4-D) great circle distance on \mathbf{Q} between \mathbf{q}_1 and \mathbf{q}_2 . We show below that d satisfies property (iv) of Section 4.2.

Lemma 3. Let \mathbf{q} be a unit length quaternion. Then for any \mathbf{q}_1 and \mathbf{q}_2

$$d(\mathbf{q}\mathbf{q}_1, \mathbf{q}\mathbf{q}_2) = d(\mathbf{q}_1, \mathbf{q}_2)$$

and

$$d(\mathbf{q}_1\mathbf{q}, \mathbf{q}_2\mathbf{q}) = d(\mathbf{q}_1, \mathbf{q}_2).$$

Proof of the Lemma 3. Let $\langle \mathbf{q}_1, \mathbf{q}_2 \rangle$ denote the dot product of \mathbf{q}_1 and \mathbf{q}_2 considered as vectors in \mathbb{R}^4 . It is easily checked that

$$\langle \mathbf{q}_1, \mathbf{q}_2 \rangle = \frac{\mathbf{q}_1 \bar{\mathbf{q}}_2 + \mathbf{q}_2 \bar{\mathbf{q}}_1}{2}.$$

Therefore

$$\begin{aligned} \langle \mathbf{q}\mathbf{q}_1, \mathbf{q}\mathbf{q}_2 \rangle &= \frac{\mathbf{q}\mathbf{q}_1 \bar{\mathbf{q}}\mathbf{q}_2 + \mathbf{q}\mathbf{q}_2 \bar{\mathbf{q}}\mathbf{q}_1}{2} = \mathbf{q} \frac{\mathbf{q}_1 \bar{\mathbf{q}}_2 + \mathbf{q}_2 \bar{\mathbf{q}}_1}{2} \bar{\mathbf{q}} \\ &= \mathbf{q} \langle \mathbf{q}_1, \mathbf{q}_2 \rangle \bar{\mathbf{q}} = \mathbf{q} \bar{\mathbf{q}} \langle \mathbf{q}_1, \mathbf{q}_2 \rangle = \langle \mathbf{q}_1, \mathbf{q}_2 \rangle \end{aligned}$$

where we have used the fact that the real number $\langle \mathbf{q}_1, \mathbf{q}_2 \rangle$ commutes with any quaternion and, that since \mathbf{q} has unit length, $\mathbf{q}\bar{\mathbf{q}} = 1$. Thus left multiplication $L(\mathbf{q})$ by \mathbf{q} preserves the dot product in \mathbb{R}^4 and *a fortiori* the distance d induced by dot product on \mathbf{Q} . The proof for right multiplication $R(\mathbf{q})$ is similar.

Proposition. If a Riemannian distance function \bar{D} on $SO(3)$ satisfies (14) and (15) above, then there is a constant c such that for all \mathbf{A} and \mathbf{B} , $\bar{D}(\mathbf{A}, \mathbf{B}) = cD(\mathbf{A}, \mathbf{B})$.

Proof of the proposition. We will consider \bar{D} as a Riemannian distance function on \mathbf{Q} defined by the Riemannian metric $\langle \cdot, \cdot \rangle$. What we need to show is that if $\langle \cdot, \cdot \rangle$ satisfies

$$\langle \mathbf{q}\mathbf{h}_1\bar{\mathbf{q}}, \mathbf{q}\mathbf{h}_2\bar{\mathbf{q}} \rangle = \langle \mathbf{h}_1, \mathbf{h}_2 \rangle \quad (\text{A15})$$

$$\langle \mathbf{h}_1\mathbf{q}, \mathbf{h}_2\mathbf{q} \rangle = \langle \mathbf{h}_1, \mathbf{h}_2 \rangle \quad (\text{A16})$$

where \mathbf{q} is in \mathbf{Q} and \mathbf{h}_1 and \mathbf{h}_2 are tangent to \mathbf{Q} at the same point, then $\langle \cdot, \cdot \rangle$ is a multiple of Euclidean dot product.

Consider first the implications of (A15). We restrict our attention to \mathbf{h}_1 and \mathbf{h}_2 pure quaternions, that is vectors tangent to \mathbf{Q} at 1. Let $c = \langle \mathbf{i}, \mathbf{i} \rangle$. If \mathbf{h} is an arbitrary unit length pure quaternion, we can find a rotation $\mathbf{A}(\mathbf{q})$ for appropriately selected \mathbf{q} so that $\mathbf{A}(\mathbf{q})\mathbf{i} = \mathbf{h}$. Then (A15) and Lemma 1 imply that $\langle \mathbf{h}, \mathbf{h} \rangle = c$. Then using \cdot to represent Euclidean dot product

$$\langle \mathbf{h}_1, \mathbf{h}_2 \rangle = \frac{\langle \mathbf{h}_1 + \mathbf{h}_2, \mathbf{h}_1 + \mathbf{h}_2 \rangle - \langle \mathbf{h}_1 - \mathbf{h}_2, \mathbf{h}_1 - \mathbf{h}_2 \rangle}{4}$$

$$= \frac{c}{4} [(\mathbf{h}_1 + \mathbf{h}_2) \cdot (\mathbf{h}_1 + \mathbf{h}_2) - (\mathbf{h}_1 - \mathbf{h}_2) \cdot (\mathbf{h}_1 - \mathbf{h}_2)]$$

$$= c\mathbf{h}_1 \cdot \mathbf{h}_2.$$

This implies that $\langle \cdot, \cdot \rangle$ is a multiple of dot product on the tangent space to \mathbf{Q} at 1.

Now if \mathbf{h}_1 and \mathbf{h}_2 are tangent to \mathbf{Q} at the quaternion \mathbf{q} , then $\mathbf{h}_1\bar{\mathbf{q}}$ and $\mathbf{h}_2\bar{\mathbf{q}}$ are tangent to \mathbf{Q} at 1 and hence (A16) implies

$$\begin{aligned}\langle \mathbf{h}_1, \mathbf{h}_2 \rangle &= \langle \mathbf{h}_1\bar{\mathbf{q}}, \mathbf{h}_2\bar{\mathbf{q}} \rangle = c(\mathbf{h}_1\bar{\mathbf{q}}) \cdot (\mathbf{h}_2\bar{\mathbf{q}}) \\ &= c[\mathbf{R}(\bar{\mathbf{q}})\mathbf{h}_1] \cdot [\mathbf{R}(\bar{\mathbf{q}})\mathbf{h}_2] = c\mathbf{h}_1 \cdot \mathbf{h}_2\end{aligned}$$

using the proof and notation of the previous lemma. The proposition follows.